Determination of Near Surface Shear-Wave Velocities in the Central Los Angeles Basin with Dense Arrays

Zhe Jia[1], Robert W Clayton[1]

[1]Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125, USA.

Corresponding author. Zhe Jia (zjia@gps.caltech.edu)

Key Points

- We derive top 3 km Vs structure of the central Los Angeles Basin using dense industrial arrays correlated with broadband seismic stations.
- Our model better resolves small scale heterogeneities and better predicts dispersion observations compared with the SCEC CVM models.
- Our model illuminates the Newport-Inglewood fault zone by a NW-SE trending high shear wave velocity belt.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2020JB021369.

Abstract

In this study, we investigate the shallow shear wave velocity structure of the Los Angeles Basin in southern California, using ambient noise correlations between 5 dense arrays and 21 broadband stations from the Southern California Seismic Network (SCSN). We observe clear fundamental mode and first overtone Rayleigh waves in the frequency band 0.25-2.0 Hz, and obtain group velocity maps through tomography. We further derive a 3D shear wave velocity model, covering a large portion of the central LA Basin for the depths shallower than 3 km. We found that the small scale shallow velocity structure heterogeneities are better resolved compared with the SCEC Community velocity models. Our model captures the presence of the Newport-Inglewood fault by a NW-SE trending high velocity belt. Our model provides more accurate constraints on local ground motion predictions with detailed mapping of structural heterogeneities.

Introduction

Los Angeles (LA) Basin is a major sedimentary basin that formed as part of the extension following the rotation of the Transverse Ranges in the Miocene. It has a depth of approximately 8 km at its deepest point and has strong lateral variations in its near-surface structure [*Shaw et al.*, 2015] due to variations of the depositional environment across the basin. Both the depth extent and near surface velocities strongly affect the amplification of ground motions. The deeper structure of the basin controls the low-frequency resonance behavior of the motion, which is important for high-rise buildings, while the shallower velocities primarily affect the motions at frequencies of 1 Hz or higher. The latter is important for community infrastructure and buildings that are less than 10-stories high. Recent studies have shown that the ground motions can vary laterally by a factor of 5 over a kilometer distance [*Clayton et al.*, 2019], which poses a significant earthquake hazard concern in the densely populated greater LA area, consequently it is important to constrain the near-surface structure and wave velocities of the LA Basin.

Studies of the detailed basin velocity structure rely on the analysis of earthquake seismic records, borehole geophysical data, geologic inference, and seismic reflection profiles. Structural models of the southern California region have also been conducted using seismic observations including receiver functions [*Ma and Clayton*, 2016; *Zhu and Kanamori*, 2000] and seismic tomography [*Hauksson*, 2000; *Qiu et al.*, 2019; *Tape et al.*, 2009]. On the other hand, Suss and Shaw [2003] measured P wave velocities of the central LA Basin through a normal moveout analysis of oil-company reflection surveys and from borehole sonic logs. All of these approaches provide constraints on the structure of the LA Basin on a variety of length scales. These various approaches have resulted in a series of 3D community velocity models (CVM) for southern California [*Lee et al.*, 2014b; *Shaw et al.*, 2015; *Small et al.*, 2017]. The CVM models are reported to be capable of forward predicting waveforms up to 0.2 Hz [*Jia et al.*, 2020; *Lee et al.*, 2014a; *Taborda et al.*, 2016]. However, the LA Basin is only covered by approximately 20 Southern California Seismic Network broadband stations, which limits the spatial resolution of the tomography based approaches. Reflection surveys and borehole logs used for velocity analysis are of

much higher precision, but they are generally only available in the oil producing areas which constitute only a small portion of the LA Basin. The detailed structure of the LA Basin is difficult to determine because it requires dense arrays of seismic instruments deployed in urban areas.

In recent years, the development of ambient noise tomography and the increasingly extensive usage of dense seismic arrays are two innovations that allow a more detailed and comprehensive structural analysis. Ambient noise surface wave tomography has been widely applied in imaging subsurface structure [*Ekström*, 2014; *Lin et al.*, 2008; *Shapiro et al.*, 2005]. Ambient noise tomography involves noise correlations, which turns coherent seismic noise recorded by different stations into Green's functions between these stations, assuming a diffuse and isotropic noise field [*Snieder*, 2004]. The dominant signal in the vertical-vertical correlations is the Rayleigh wave. The next steps include measuring dispersions of surface waves with a frequency-time analysis [*Bensen et al.*, 2007; *Levshin et al.*, 1972], standard seismic tomography [*Barmin et al.*, 2001] and nonlinear shear wave inversions [*Herrmann*, 2013]. Ambient noise tomography has been traditionally used to investigate regional and continental scale structure with long period (T>5s) surface waves obtained from correlations between broadband seismic stations. Recent developments in dense array deployments extend the ambient-noise based structural studies to higher frequencies, making seismic imaging of local and shallow velocity structure easily applicable [*Castellanos et al.*, 2020; *Lin et al.*, 2013; *Roux et al.*, 2016].

Five temporary dense arrays of totaled more than 16000 high-frequency velocity sensors were deployed by oil companies from 2011 to 2018 in the Los Angeles region (Fig. 1). Most of these arrays are located above oil fields associated with the Newport-Inglewood fault. An exception is the Santa Fe Springs survey, which is closer to the northeastern edge of the basin. The stations are deployed for different time spans from 3 weeks to 5 months and recorded continuously during those times. This means they not only record the active source signals of the survey but also the passive ambient noise. Although the dense nodal arrays only cover small areas, and each of them was deployed at different times, they all overlap in time with the long-term SCSN broadband seismic stations which span over the whole LA Basin. This

long span time and dense coverage of the inter-array ambient noise correlations provide a unique opportunity to investigate the detailed shallow structural heterogeneities of the central LA Basin.

In this study, we will illustrate the effectiveness of noise correlations between the broadband SCSN stations and the high-frequency dense array nodal sensors between 0.2-2.0 Hz. We then perform frequency-time analysis on the noise correlation functions to retrieve Rayleigh-wave dispersion curves. Using a standard straight-ray tomography approach, we generate group velocity maps, which are then inverted for the shear-wave velocity structure. We then assess the model errors, and discuss the effect of our model on strong ground motion predictions.

Data and method

We collect the ambient noise data recorded by the industrial arrays (16,000 short-period vertical component nodes) and SCSN stations (21 broadband sensors) in their overlapping time periods. This results in at least 500 hours of noise data for each of the node-broadband pairs (Fig. 1). The industry data is recorded with dense arrays (100 stations/km^2) in which each station is a vertical-component velocity sensor with a corner frequency of 10 Hz. These geophones are deployed in a relatively noisy urban setting, and although they are most sensitive to frequencies above 10 Hz, several studies have shown that noise correlations within the dense arrays can produce Rayleigh waves in the frequency band of 0.1~10 Hz [*Ben-Zion et al.*, 2015; *Lin et al.*, 2013; *Nakata et al.*, 2015].

We follow Bensen et al. [2007] to calculate the inter-array noise correlation functions (NCFs). We cut the vertical component noise data of both SCSN stations and dense arrays into one hour segments, and then perform time domain normalization and spectral whitening to suppress the influence of earthquakes and other coherent signals. Due to the reduced instrumental responses at lower frequencies for the dense array geophones, we normalize the spectrum of geophone noise data by their spectral amplitudes. That is, we force the amplitudes of all frequencies to be uniform and only keep their phase information. We then

calculate the NCF of all possible inter-array station pairs, and stack these 1-hour correlations over a three-week span. We have computed 353006 noise correlations, including 106800 for array LB3D, 104811 for SB3D, 50660 for SFS, 48840 for ELB, and 41895 for ROSE. We summarize these numbers in Table S1. Finally, we average both sides of each NCF by flipping the virtual source and receiver, forming symmetric waves on the positive side and negative side [*Lin et al.*, 2008]. In addition to the inter-array correlations, we also calculate the intra-array correlations within the SCSN stations. An example of the distance-sorted stacked NCFs is shown in Fig. 2. For a relatively broad frequency band between 0.2-2.0 Hz (Fig. 2a), we observe a clear trend of the Rayleigh wave moveout but it is difficult to distinguish the separate modes. However, for relatively higher frequencies between 0.5-2.0 Hz (Fig. 2b), we observe splitting of the modes, indicating that both fundamental and first overtone Rayleigh waves exist in the NCFs. The propagations of these multi-modal Rayleigh waves are shown in Fig. 3, as wavefields recorded by four dense arrays using the SCSN station LTP as a virtual source. We observe both slower waves with large amplitudes and faster but weaker signals for the four different directions. These are the fundamental mode and first overtone Rayleigh waves, respectively. Note that the wave propagation within every single dense array involves nontrivial phase distortions, amplitude variations, and frequency content differences. Presumably, these different paths could have very different small scale scatter distributions. Another possible cause is the difference of intrinsic attenuation effect along these ray paths. More intensely scattered high frequency waves and less intrinsic attenuation can make the surface wavefield more complex. The observed wavefield complexities suggest significant structural heterogeneities along the ray paths.

We use a frequency-time analysis (FTAN) method [*Herrmann*, 2013; *Levshin et al.*, 1972] to retrieve both the fundamental and first overtone modes of Rayleigh wave group velocity dispersions for all SCSN-array pairs. As for the data selection, we empirically set the minimum signal-to-noise ratio to be 5. We filter the waves between 0.2~1.0 Hz before we screen them with SNR. For the NCF with a distance l, we define the signal time window between $l/1.2$~$l/0.3$ s, and the noise window to be 0~$l/1.2$ s, where 1.2 and

0.3 km/s are proxies of upper and lower bounds of Rayleigh waves velocities. Both signal and noise are measured by the root mean square value of the waves over time. We only keep the interstation distance of the NCFs greater than 8 km. The distance restriction is to allow adequate separation of the modes. Furthermore, to ensure the robustness of the dispersion measurements, we stack the NCFs for almost identical ray paths defined by dense array bins of 300-meter radius, typically involves tens of stations. There are reasons why we choose the circular distance to be 300 meters. The first is to avoid gaps in the arrays caused by infrastructure. Secondly, our correlation distance in tomography is 1km, and a larger stacking radius is closer to this wavelength limit, which could potentially introduce artificial smoothness. Furthermore, to test the nodal array interstation distance needed in this study, we down sampled the nodal station density to average interstation distances to 0.3, 0.45 and 0.7 km, and we conducted checkboard tests at period of 2s in these cases. We found that an average interstation distance of less than 0.3km is required for recovering the checkboard map (Fig. S1). In the FTAN approach, we apply a Gaussian filter where we set the filter parameter alpha to 25 as a compromise between the narrow-band assumption and filtering robustness [*Zhang et al.*, 2020]. Once we obtained the velocity-time amplitudes, we adopt a hybrid dispersion picking strategy, in which we first automatically pick the dispersion curves within velocity-time corridors, and then manually check the picked curves. The corridors are estimated through stacking the velocity-time amplitudes of NCFs from each SCSN station to all geophones in each dense array, because these NCFs share similar ray paths. Fig. S2a shows the example of corridors on the stacked FTAN maps from the virtual source LGB to the different dense arrays. The fundamental mode and first overtone surface wave signals are well covered by the corridors. Moreover, the picked dispersions for LGB against array LB3D (Fig. S2a) is consistent with machine learning picked ones [*Zhang et al.*, 2020]. Fig. S2b presents an overview of all corridors on the stacked velocity-time maps. Within the large number of noise correlations, we manually check one dispersion curve at equal intervals every 50 curves. Because each dense array has thousands of nodal stations, and the ray paths from one virtual source to nodal stations of same array are similar to each other, this sampling inspection approach provide robust way of assessing the quality of corridors and the criteria used for accepting the picks. We accept the picks within

the corridor when their spectral amplitudes are larger than all other possible picks at the same frequency, and when their spectral amplitudes are larger than 0.4 cm-s for the fundamental mode and 0.3 cm-s for the first overtone. These thresholds provided good quality picks with a reasonably high acceptance rate. Fig. 4 shows an example of the group velocity dispersion picking for the pair of broadband station LTP and dense array geophone 1041-5011. The dispersion curves of the fundamental mode and first overtone Rayleigh waves are clear and are well picked.

Due to the uneven data quality between each station-array pair, the numbers of picked dispersion data for different SCSN station and nodal array pairs can be significantly different (Fig. S3). If their weights are not balanced, the tomographic inversion will be dominated by those station-array pairs that contribute most dispersion picks. Therefore, we adjust the weightings of different station-array pairs by uniformly resampling so no more than 200 picks for each station-array pair and each period are used. The dispersion counts from different station-array pairs after this reweighting are much more even (Fig. S3), indicating that the data are better weighted. A compilation of all dispersion data after reweighting is shown in Fig. 5. These curves show an overall increase of group velocity U with period, and hence an increase of Vs with depth. Fig. 6 shows the compilation of extracted fundamental mode and first overtone Rayleigh wave group velocities for periods of 1s, 2s, 3s and 4s.

To analyze the inter-station group velocity dispersion measurements, we use the 2D straight-ray tomography adapted from Barmin et al. [2001] to derive the fundamental mode and first overtone Rayleigh wave group velocity maps. The region is discretized into $0.012^\circ \times 0.008^\circ$ grids, and we linearly invert for group velocity maps of each period between 0.5-4.0 s. In the linear inversions, we adopt two penalty terms for regularization, 1) constraint on model smoothness; 2) damping as a function of the ray coverage density. For the smoothness constraints, we set the correlation length to be 1 km. The ray coverage damping uses the concept of azimuthal coverage proposed by Ekström [2006] to quantitatively indicate the effective ray coverage in each grid. We perform this linear inversion with two iterations. In the first iteration, we weight all the dispersion data equally, and apply a strong regularization

for a smooth model which presumably represents the first order structural heterogeneities. We then forward predict the synthetic group arrival times for all ray paths, and approximate the misfit between data and synthetic arrival times as data error. We discard those data with misfits are larger than three times of standard deviation [*Castellanos et al.*, 2018]. In the second iteration, we construct the data covariance matrix which diagonal terms are specified based on the data errors. Thus, the data weightings are inversely proportional to the data error, making the group velocity map inversions less dependent on specific paths with large misfits. To understand the data resolvability in our group velocity map inversions, we generate standard checkboard resolution maps through the resolution matrix $(G^tC^{-1}G+Q)^{-1}G^tC^{-1}Gm^*$ [*Ma and Clayton*, 2014], where C is the data covariance matrix, Q is the regularization matrix, G is the forward operator, and m* is the input checkboards containing up to $\pm$20% contrasting slowness variations. Fig. 7 shows the checkboard resolution maps at periods of 1, 2s, 3s and 4s for both fundamental mode and first overtone Rayleigh waves. We found that the contrasting checkboard structures are well recovered in a large portion of the LA Basin for all these periods and wave modes. However, the resolution is lower at the edge of the studied area due to the poor azimuthal coverage of the rays.

We invert for the shear wave velocities as a function of depth, using 1D group velocity profiles at each grid extracted from the tomographic maps. We use the SURF96 program [*Herrmann*, 2013] for this shear wave velocity inversion. SURF96 is designed for horizontally isotropic layered structure, and involves an iterative least squares inversion procedure. Using the SCEC CVM-S4 profiles embedded with Vs30-derived geotechnical layer [*Ely et al.*, 2010] at all grid points as the initial model, we iteratively perturb the shear wave velocities of all layers while calculating P wave velocities following the empirical relationship introduced by Brocher [2005], until a good fit between observed and synthetic fundamental mode and first overtone Rayleigh wave dispersion curves are achieved. These final 1D shear wave velocity profiles are integrated into a whole 3D Vs model of the study region. We also tested using the 1D

model inverted from the mean of all dispersion curves as the initial model, and found no significant change to the shear wave velocity profiles.

Results

We invert the extracted dispersion curves from the Rayleigh waves group velocity maps for a suite of frequencies between 0.25-2.0 Hz. Fig. 8 shows the fundamental mode and first overtone group velocities at periods of 1s, 2s, 3s and 4s. For both modes and all periods, we observed two high group velocity anomalies in the southwestern and northeastern corners of the study region, which correspond to the Palos Verdes Hills and Puente Hills, respectively. The group velocities are much lower in the area between these high velocity hills, which suggests the decreasing sedimentary layer thickness towards the Palos Verde fault zone to the southwest and towards the Elsinore fault zone to the northeast, consistent with the hard bedrocks beneath these hills that bound the LA Basin in these two directions. Note that the well resolved area is smaller for the short-period and long-period end (period of 1s and 4s in Fig. 7). The reason why the areas shrink at these two period ends is that the surface wave energy decreases when period approaches 0.5s and 5s (e.g. Fig. 4b). The depleted noise correlation function energy for the short period end is primarily due to the low signal-to-noise ratio of high frequency Rayleigh waves. At the long period end, the lower instrumental responses of the high frequency industrial nodal sensors cause the noise signals less correlated. In the center of the LA Basin, we observe significant group velocity variations of smaller length scales. These variations are far beyond the group velocity errors (Fig. S4) estimated from the diagonal term of the model covariance matrix $C=(G^tC^{-1}G+Q)^{-1}$ [*Ma and Clayton*, 2014]. Such small scale variations indicate nontrivial structural heterogeneity of the sedimentary layer, and suggest possible interface roughness within the basin. The effectiveness of data inversion is shown by the significant slowness variance reductions for different modes and periods (Fig. S5).

We show the shear wave velocity inversions in Fig. 9. Fig. 9a shows horizontal slices of the shear wave velocities at depths of 0.5-2.5 km. At larger depths, the Vs velocity continues to increase and the low velocity area shrinks. We estimate the Vs uncertainties for each grid (e.g. Fig. S6) by inverting 50

dispersion curves sampled from group velocity errors. Specifically, for each location grid, we generate 50 randomly perturbed dispersion curves for fundamental mode and first overtone, respectively. At an individual period for each mode, the group velocities are perturbed following a Gaussian distribution with mean of the inverted group velocities (Fig. 8) and with standard deviations from the diagonal term of the model covariance matrix in the tomographic inversions (Fig. S4). After we obtained the ensemble of dispersion curves (e.g. Fig. S6a), we invert them for an ensemble of 1D shear wave velocity models (e.g. Fig. S6b). We further measure the standard deviation error of Vs at each depth using the ensemble of 1D Vs models, and combine these 1D Vs errors to Vs error maps (Fig. S7). The measured Vs standard deviation error for most of the central LA Basin (Fig. S7) is significantly smaller than the observed Vs variations in our model (Fig. 9). One pronounced feature of the Vs model is the high velocity belt from NW to SE at the depths of around 1.5 km (Fig. 9a). This high velocity strand, accompanied with a sharp horizontal velocity contrast (Fig. 9b), appears to be consistent with the trace of the Newport-Inglewood fault system (Fig. 9a), which is manifested by a line of hills from Cheviot Hills on the northwestern end to the Signal Hill on the southeastern end [*Hauksson*, 1987]. These hills are the topographic expression of the Newport-Inglewood fault extending to the surface, suggesting that the fault movement transported higher velocity materials from deeper to shallower depths. This explains the observed high shear wave velocity anomalies of the Newport-Inglewood fault zone and is consistent with various dense array studies [*Bianco et al.*, 2019; *Castellanos et al.*, 2020; *Lin et al.*, 2013; *Nakata et al.*, 2015].

We note in the Vs slices at a depth of 1-1.5 km (Fig 9) that there are low-velocity streaks trending NW to SE, to the northeast of the aforementioned NW-SE high velocity belt. This NW-SE trending low velocity feature corresponds well with the P wave low velocity anomalies obtained with sonic logs and industrial reflection data [*Süss and Shaw*, 2003]. It also agrees with the trend of sedimentary deposits during late Miocene [*Redin*, 1991], suggesting that they may be due to depositional channels created when the LA Basin was filling with sediments in the Miocene [*Süss and Shaw*, 2003].

Discussion

We compare our shear wave velocity model with SCEC CVM-S4 and CVM-H15.1 models as shown in Fig. 10. These models generally agree on the large scale structure of the basin [*Taborda et al.*, 2016]. However, their near-surface structures of the LA Basin appear different. The CVM-S model is smoother since it is constructed using full-waveform tomography. On the other hand, since the CVM-H model is largely inferred from geological observations and P wave velocities from oil industry, it has geotechnical layers with sharp artificial boundaries. In the shallow basin, the CVM-S model better characterizes the first-order velocity structure, while the CVM-H model captures finer scale heterogeneities. A comparison of recorded and synthetic waveforms in Southern California confirms that the CVM-S model predicts long period waveforms better, while the CVM-H model better fits the arrival time of scattered waves [*Lai et al.*, 2020]. Our Vs model of the LA Basin, named LAS1 (Los Angeles Shallow 1), is in between the CVM-S and CVM-H models, as its overall basin shape and velocity range is closer to the CVM-S model, while rapid lateral variations such as the high velocity anomaly of the Newport-Inglewood fault are similar to the CVM-H model (Fig. 10a).

The small scale shallow structural variation of our LAS1 model appears different from both the SCEC CVM-S and CVM-H models, which may significantly influence the modeling of earthquake ground motion amplitudes for the central LA Basin. To illustrate the effect of the variations in the near surface velocities, we performed 2D simulations of vertically-incidental shear wave pulse using the CVM-S, CVM-H, and our LAS1 models for each cross-section. The calculations are done with a 2D fourth-order staggered-grid finite-difference formulation [*Virieux*, 1984] that initiated a 1 Hz SH-pulse (in velocity) at a depth of 3 km and propagated it to the surface (Fig. S8). The peak acceleration, obtained through numerical differentiation at each horizontal position, is then compared to an equivalent pulse propagated through a medium with a constant shear velocity of 1 km/s (Fig. S8). The ratios of the results are displayed in Figure 10b and show that the LAS1 and CVM-H models have comparable levels of peak acceleration and lateral variations, but that the CVM-S has significantly lower motions and is laterally

smoother. This is likely due to the different wavelengths that these Vs models feature. The CVM-S model comes from a long period (T>5s) full-waveform inversion, while our LAS1 model captures shorter wavelength structural features from short period data (0.5-4s), and the CVM-H model incorporates high-frequency information from geotechnical layers and seismic reflection constraints. The small wavelength heterogeneities can produce larger local lateral and vertical velocity gradient, which could significantly amplify the strong motions [*Bowden et al.*, 2015; *Steidl et al.*, 1996]. Additionally, scattering and diffraction due to the existence of structural heterogeneities and rough basin interfaces can cause wave interference, and potentially amplify the motions [*Delépine and Semblat*, 2012]. With the additional constraints from dense array dispersion data, our model can potentially predict the variations of shallow crustal amplifications in the LA Basin which is observed during the 2019 Ridgecrest earthquake sequence [*Filippizitis et al.*, 2021]. The simulations shown here are only intended to illustrate the differences in the models and we present them in lieu of a more complicated 3D simulation.

To better understand the Vs structural difference between our LAS1 model and the CVM-S/CVM-H models, we forward calculate the fundamental model and first overtone Rayleigh wave group velocity dispersions for these models. We used the program sdisp96 [*Herrmann*, 2013] to calculate dispersion curves for multiple 1D profiles and then combined them to 2D maps at different depths. The comparison of these synthetic dispersion maps against the observed tomographic group velocity maps is shown in Fig. 11. Among these models, the LAS1 group velocity maps fit the tomographic observations the best. The CVM-S model over predicts the long period group velocities, while the CVM-H model produces much lower Rayleigh group velocities than the observations. To directly examine the model predictions against the picked dispersion data, we generate synthetic dispersion curves from group velocity maps of our tomographic result and the CVM-S/CVM-H models along the ray paths between SCSN stations and dense arrays. The fittings to directly picked dispersion data for these models are shown in Fig. 12. Similar to the group velocity map comparisons, the CVM-S model predicts that long period Rayleigh waves

travel faster than observations, and the CVM-H model underestimates the Rayleigh wave group velocities for multiple frequency bands. Our model fits the observations significantly better than CVM-S/CVM-H models, but still generate non-negligible misfits. Possible sources of the misfits are the simple forward modeling used in the problem, in particular the straight ray assumption, and the uneven ray coverage. The diminishing ray density toward the edge of the basin contributes to the data misfit. A possible solution to this is to use machine-learning based tomography to reduce the artifacts [*Bianco et al.*, 2019].

Conclusion

We use ambient noise correlations between 5 dense oil company nodal networks and 21 broadband Southern California Seismic Network stations to develop a high resolution shallow structural velocity model of the central Los Angeles Basin. We demonstrated that this short period (0.25-2.0 Hz) inter-station noise correlation can produce clear multimodal Rayleigh wavefields, thus can provide very detailed coverage of a large portion of the LA Basin. We extracted group velocity dispersion curves and inverted for the shear wave velocity structure. Our Vs model shows the shrinking circumference of the velocity sedimentary materials as the basin deepens, and reveals the existence of multiscale structural heterogeneities and interface roughness within the basin. The model also shows evidence of the Newport-Inglewood fault as a high shear-wave velocity belt. The estimated Vs model generally agrees with the SCEC CVM-S and CVM-H models, but better fits the Rayleigh wave dispersion data. It predicts shallow crustal amplification more similar to the CVM-H model than the CVM-S. This model should be able to better predict the variations in the shallow crustal amplification, which will improve strong ground motion simulations.

Acknowledgments

We gratefully thank reviewer Yixiao Sheng, an anonymous reviewer, associate editor Nori Nakata and editor Yehuda Ben-Zion for their constructive comments and suggestions. We gratefully thank Signal Hill Petroleum for providing permission to use the ambient noise data for the Long Beach (LB3D), Extended Long Beach (ELB) arrays, Breitburn Energy for the use of the Santa Fe Springs and Rosecrans arrays, and LA Seismic for providing the Seal Beach array. Access to the raw data requires permission from these owners. The Rayleigh wave group velocity dispersion data used to generate the final velocity model can be downloaded from: http://doi.org/10.22002/D1.1670. The final LAS1 wave velocity model can also be downloaded from that site. We thank the Southern California Earthquake Data Center for providing public access to the broadband seismic ambient noise data (https://scedc.caltech.edu). We thank Jorge C. Castellanos for helpful discussions. This project is supported by NSF/EAR-15200081.

References

Barmin, M., M. Ritzwoller, and A. Levshin (2001), A fast and reliable method for surface wave tomography, in *Monitoring the comprehensive nuclear-test-ban treaty: Surface waves*, edited, pp. 1351-1375, Springer.

Ben-Zion, Y., F. L. Vernon, Y. Ozakin, D. Zigone, Z. E. Ross, H. Meng, M. White, J. Reyes, D. Hollis, and M. Barklage (2015), Basic data features and results from a spatially dense seismic array on the San Jacinto fault zone, *Geophysical Journal International*, *202*(1), 370-380.

Bensen, G., M. Ritzwoller, M. Barmin, A. L. Levshin, F. Lin, M. Moschetti, N. Shapiro, and Y. Yang (2007), Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophysical Journal International*, *169*(3), 1239-1260.

Bianco, M. J., P. Gerstoft, K. B. Olsen, and F.-C. Lin (2019), High-resolution seismic tomography of Long Beach, CA using machine learning, *Scientific reports*, *9*(1), 1-11.

Bowden, D., V. C. Tsai, and F. C. Lin (2015), Site amplification, attenuation, and scattering from noise correlation amplitudes across a dense array in Long Beach, CA, *Geophys. Res. Lett.*, *42*(5), 1360-1367.

Brocher, T. M. (2005), Empirical relations between elastic wavespeeds and density in the Earth's crust, *Bull. Seismol. Soc. Am.*, *95*(6), 2081-2092.

Castellanos, J. C., R. W. Clayton, and A. Juarez (2020), Using a Time- based Subarray Method to Extract and Invert Noise- derived Body Waves at Long Beach, California, *Journal of Geophysical Research: Solid Earth*, *125*(5), e2019JB018855.

Castellanos, J. C., R. W. Clayton, and X. Pérez- Campos (2018), Imaging the eastern Trans- Mexican volcanic belt with ambient seismic noise: Evidence for a slab tear, *Journal of Geophysical Research: Solid Earth*, *123*(9), 7741-7759.

Clayton, R., P. Persaud, M. Denolle, and J. Polet (2019), Exposing Los Angeles's shaky geologic underbelly, Eos, 100, edited.

Delépine, N., and J.-F. Semblat (2012), Site effects in an alpine valley with strong velocity gradient: interest and limitations of the 'classical'BEM, *Soil Dynamics and Earthquake Engineering*, *38*, 15-24.

Ekström, G. (2014), Love and Rayleigh phase-velocity maps, 5–40 s, of the western and central USA from USArray data, *Earth and Planetary Science Letters*, *402*, 42-49.

Ekström, G. r. (2006), Global detection and location of seismic sources by using surface waves, *Bulletin of the Seismological Society of America*, *96*(4A), 1201-1212.

Ely, G. P., T. Jordan, P. Small, and P. J. Maechling (2010), A VS30- derived nearsurface seismic velocity model, paper presented at Abstract S51A-1907, Fall Meeting, AGU San Francisco, CA.

Filippitzis, F., M. Kohler, T. Heaton, R. Graves, R. Clayton,R. Guy, J. Bunn, and K. Chandy, (2020), Site Response in Urban Los Angeles From the 2019 Ridgecrest Earthquake Sequence, accepted in *Earthquake Spectra*, 2020.

Hauksson, E. (1987), Seismotectonics of the Newport-Inglewood fault zone in the Los Angeles basin, southern California, *Bulletin of the Seismological Society of America*, *77*(2), 539-561.

Hauksson, E. (2000), Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in southern California, *Journal of Geophysical Research: Solid Earth*, *105*(B6), 13875-13903.

Herrmann, R. B. (2013), Computer programs in seismology: An evolving tool for instruction and research, *Seismological Research Letters*, *84*(6), 1081-1088.

Jia, Z., X. Wang, and Z. Zhan (2020), Multifault Models of the 2019 Ridgecrest Sequence Highlight Complementary Slip and Fault Junction Instability, *Geophys. Res. Lett.*, *47*(17), e2020GL089802.

Lai, V. H., R. W. Graves, C. Yu, Z. Zhan, and D. V. Helmberger (2020), Shallow Basin Structure and Attenuation are Key to Predicting Long Shaking Duration in Los Angeles Basin, *Journal of Geophysical Research: Solid Earth*, e2020JB019663.

Lee, E. J., P. Chen, and T. H. Jordan (2014a), Testing waveform predictions of 3D velocity models against two recent Los Angeles earthquakes, *Seismological Research Letters*, *85*(6), 1275-1284.

Lee, E. J., P. Chen, T. H. Jordan, P. B. Maechling, M. A. Denolle, and G. C. Beroza (2014b), Full- 3- D tomography for crustal structure in southern California based on the scattering- integral and the adjoint-wavefield methods, *Journal of Geophysical Research: Solid Earth*, *119*(8), 6421-6451.

Levshin, A. L., V. Pisarenko, and G. Pogrebinsky (1972), On a frequency-time analysis of oscillations, paper presented at Annales de geophysique, Centre National de la Recherche Scientifique.

Lin, F.-C., D. Li, R. W. Clayton, and D. Hollis (2013), High-resolution 3D shallow crustal structure in Long Beach, California: Application of ambient noise tomography on a dense seismic arrayNoise tomography with a dense array, *Geophysics*, *78*(4), Q45-Q56.

Lin, F.-C., M. P. Moschetti, and M. H. Ritzwoller (2008), Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps, *Geophysical Journal International*, *173*(1), 281-298.

Ma, Y., and R. W. Clayton (2014), The crust and uppermost mantle structure of Southern Peru from ambient noise and earthquake surface wave analysis, *Earth and Planetary Science Letters*, *395*, 61-70.

Ma, Y., and R. W. Clayton (2016), Structure of the Los Angeles Basin from ambient noise and receiver functions, *Geophysical Journal International*, *206*(3), 1645-1651.

Nakata, N., J. P. Chang, J. F. Lawrence, and P. Boué (2015), Body wave extraction and tomography at Long Beach, California, with ambient- noise interferometry, *Journal of Geophysical Research: Solid Earth*, *120*(2), 1159-1173.

Qiu, H., F. C. Lin, and Y. Ben- Zion (2019), Eikonal tomography of the Southern California plate boundary region, *Journal of Geophysical Research: Solid Earth*, *124*(9), 9755-9779.

Redin, T. (1991), Oil and Gas Production from Submarine Fans of the Los Angeles Basin: Chapter 8.

Roux, P., L. Moreau, A. Lecointre, G. Hillers, M. Campillo, Y. Ben-Zion, D. Zigone, and F. Vernon (2016), A methodological approach towards high-resolution surface wave imaging of the San Jacinto Fault Zone using ambient-noise recordings at a spatially dense array, *Geophysical Journal International*, *206*(2), 980-992.

Shapiro, N. M., M. Campillo, L. Stehly, and M. H. Ritzwoller (2005), High-resolution surface-wave tomography from ambient seismic noise, *Science*, *307*(5715), 1615-1618.

Shaw, J. H., A. Plesch, C. Tape, M. P. Suess, T. H. Jordan, G. Ely, E. Hauksson, J. Tromp, T. Tanimoto, and R. Graves (2015), Unified structural representation of the southern California crust and upper mantle, *Earth and Planetary Science Letters*, *415*, 1-15.

Small, P., D. Gill, P. J. Maechling, R. Taborda, S. Callaghan, T. H. Jordan, K. B. Olsen, G. P. Ely, and C. Goulet (2017), The SCEC unified community velocity model software framework, *Seismological Research Letters*, *88*(6), 1539-1552.

Snieder, R. (2004), Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase, *Physical Review E*, *69*(4), 046610.

Steidl, J. H., A. G. Tumarkin, and R. J. Archuleta (1996), What is a reference site?, *Bull. Seismol. Soc. Am.*, *86*(6), 1733-1748.

Süss, M. P., and J. H. Shaw (2003), P wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California, *Journal of Geophysical Research: Solid Earth*, *108*(B3).

Taborda, R., S. Azizzadeh-Roodpish, N. Khoshnevis, and K. Cheng (2016), Evaluation of the southern California seismic velocity models through simulation of recorded events, *Geophysical Journal International*, *205*(3), 1342-1364.

Tape, C., Q. Liu, A. Maggi, and J. Tromp (2009), Adjoint tomography of the southern California crust, *Science*, *325*(5943), 988-992.

Virieux, J. (1984), SH-wave propagation in heterogeneous media: Velocity-stress finite-difference method, *Geop*, *49*(11), 1933-1942.

Zhang, X., Z. Jia, Z. E. Ross, and R. W. Clayton (2020), Extracting dispersion curves from ambient noise correlations using deep learning, *IEEE Transactions on Geoscience and Remote Sensing*.

Zhu, L., and H. Kanamori (2000), Moho depth variation in southern California from teleseismic receiver functions, *Journal of Geophysical Research: Solid Earth*, *105*(B2), 2969-2980.

Figures

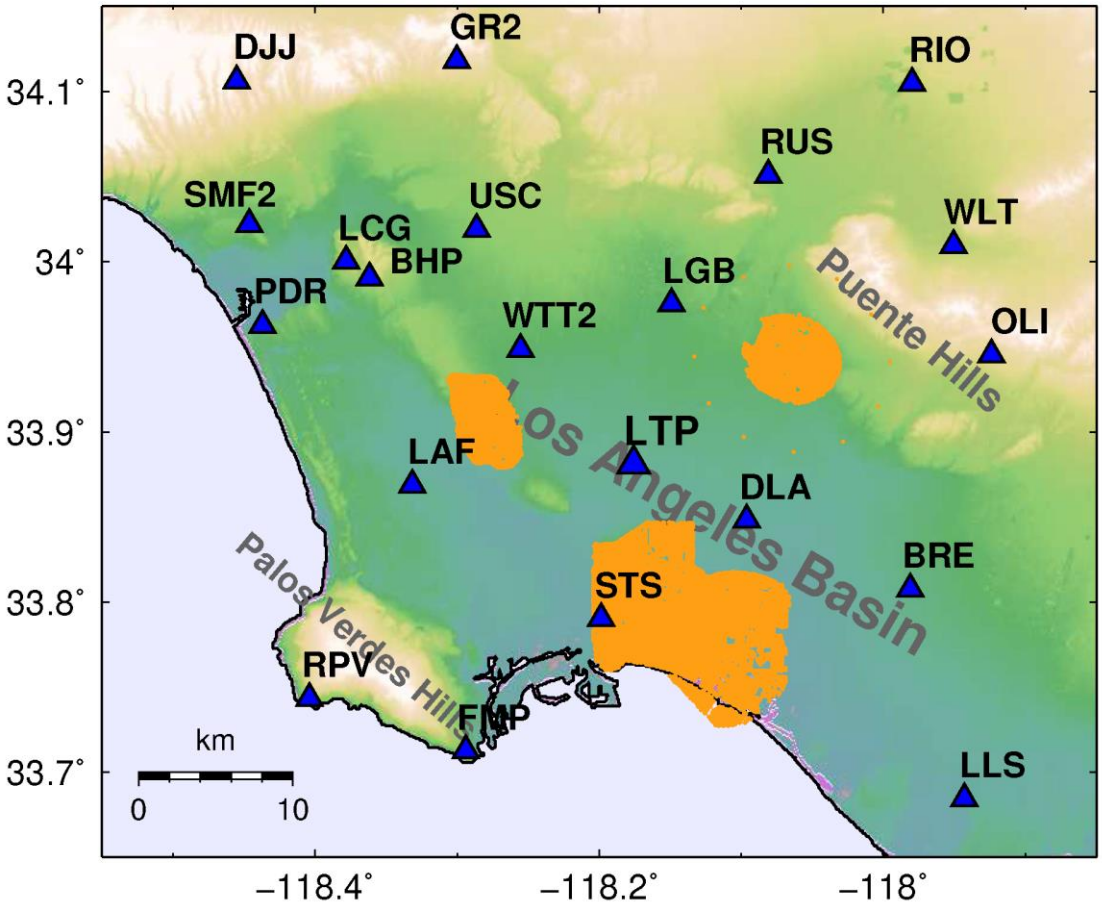


Figure 1. Distribution of the 21 broadband seismic stations (blue triangles) and 5 dense oil company nodal arrays (orange shaded areas) used in this study, the largest of which has 3 side-by-side surveys.

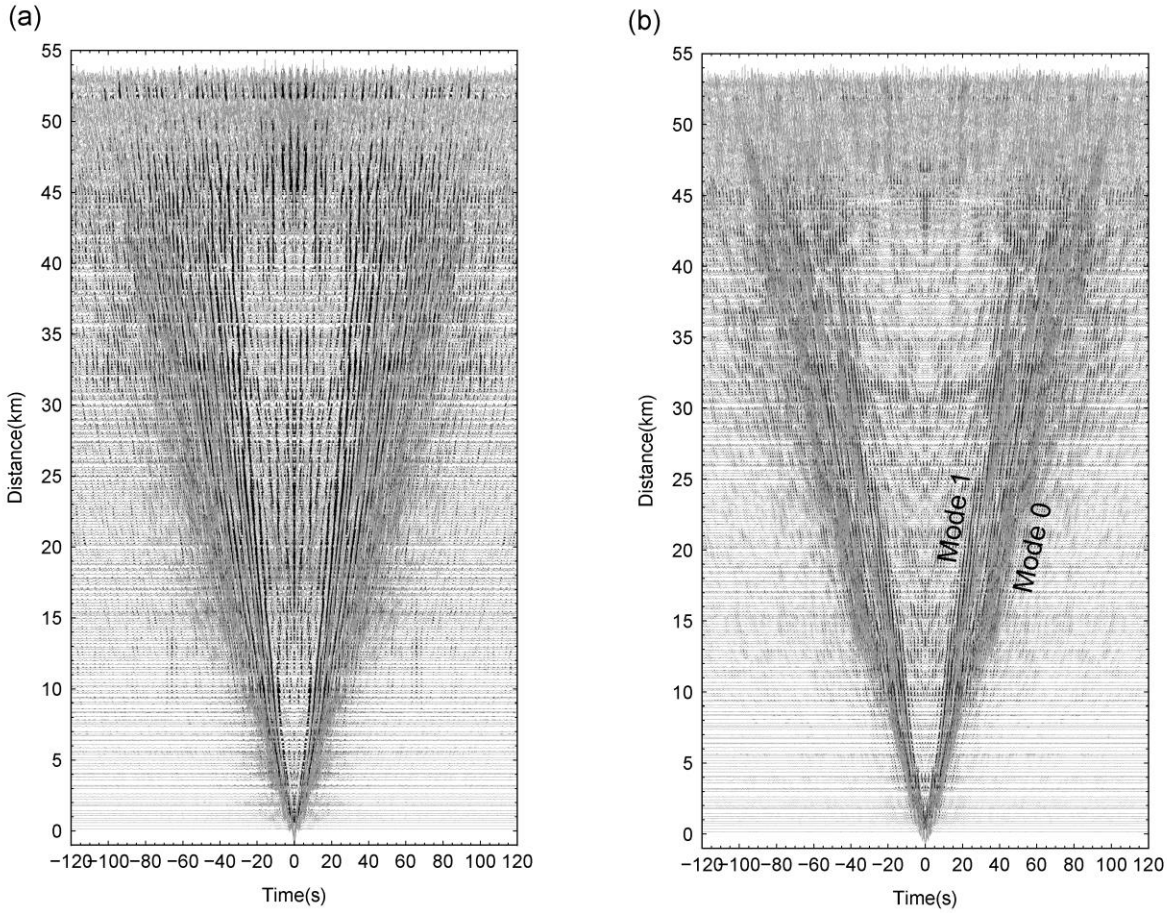


Figure 2. Inter-array noise correlation functions between SCSN stations and dense nodal arrays. (a) Noise correlation functions stacked in distance bins between 0.2-2.0 Hz. (b) High frequency (0.5-2.0 Hz) filtered noise correlation functions stacked in distance bins.

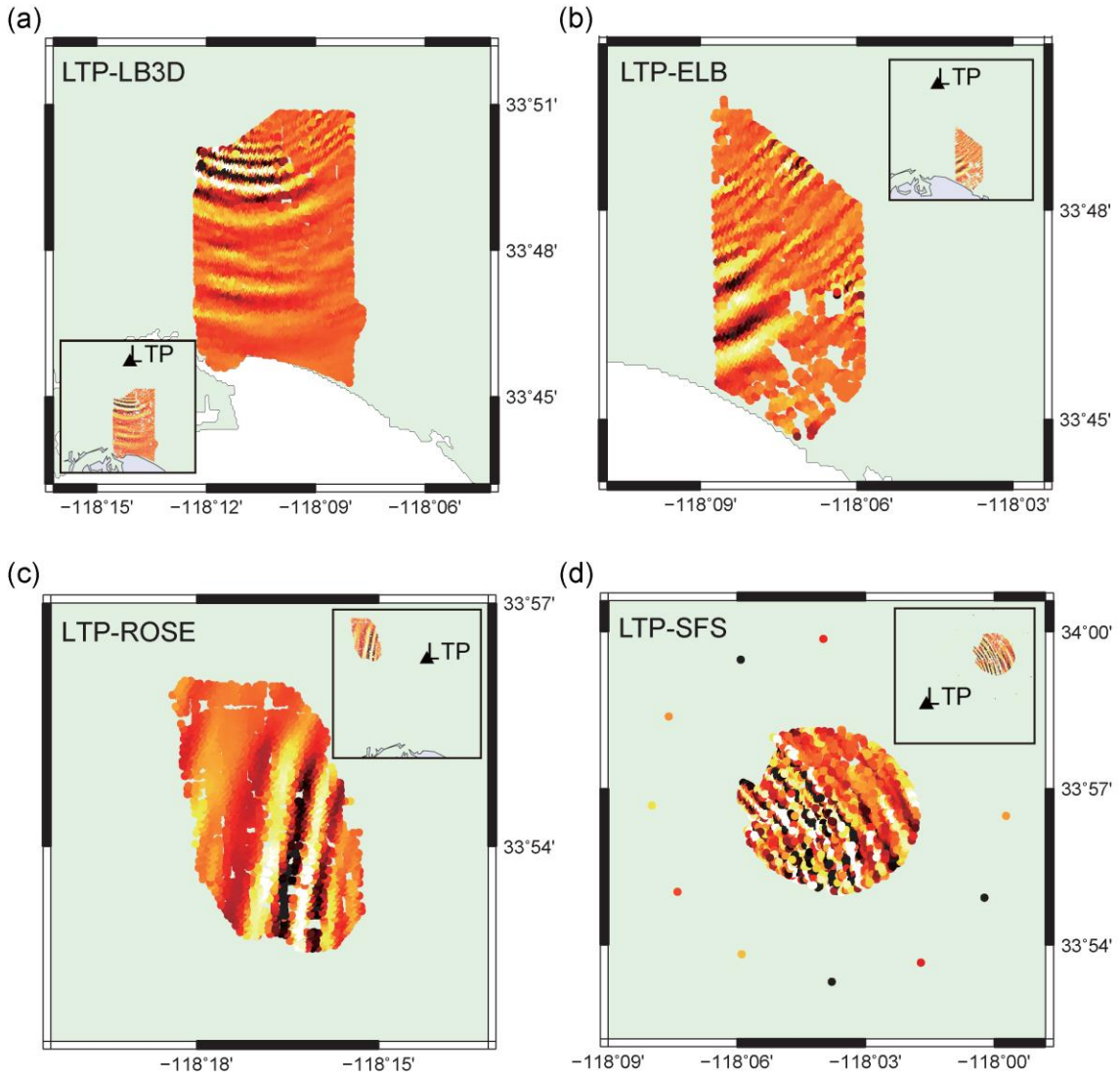


Figure 3. Wavefield snapshots of the noise correlation functions from station LTP to dense arrays (a) LB3D, (b) ELB, (c) ROSE and (d) SFS.

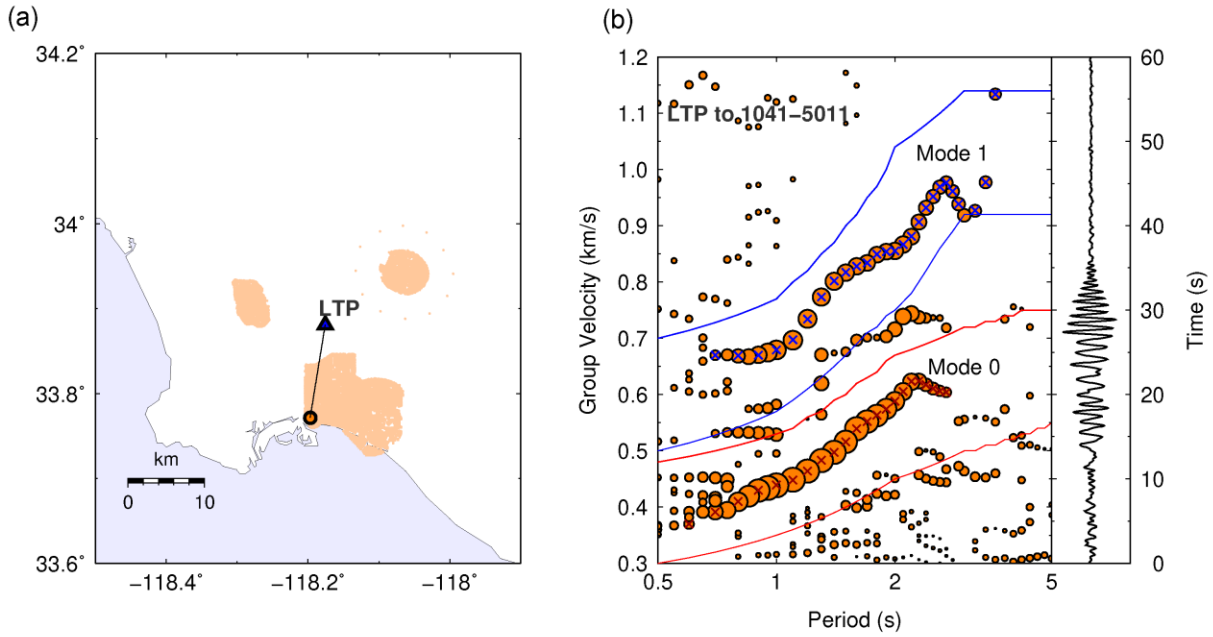


Figure 4. Example of labeling of dispersion curves. (a) The ray path from broadband station LTP (black triangle) to a dense array station 1041-5011 (orange circle). (b) The FTAN image (period axis uses a logarithmic scale) of the noise correlation function for the ray path in (a). The size of each circle is proportional to the wave packet energy at the corresponding period and group velocity. The red and blue crosses are the picked dispersion data for the fundamental mode and first overtone, respectively. The red and blue lines indicate the corridors for these two modes. The waveform to the right indicates the time-domain noise correlation function.

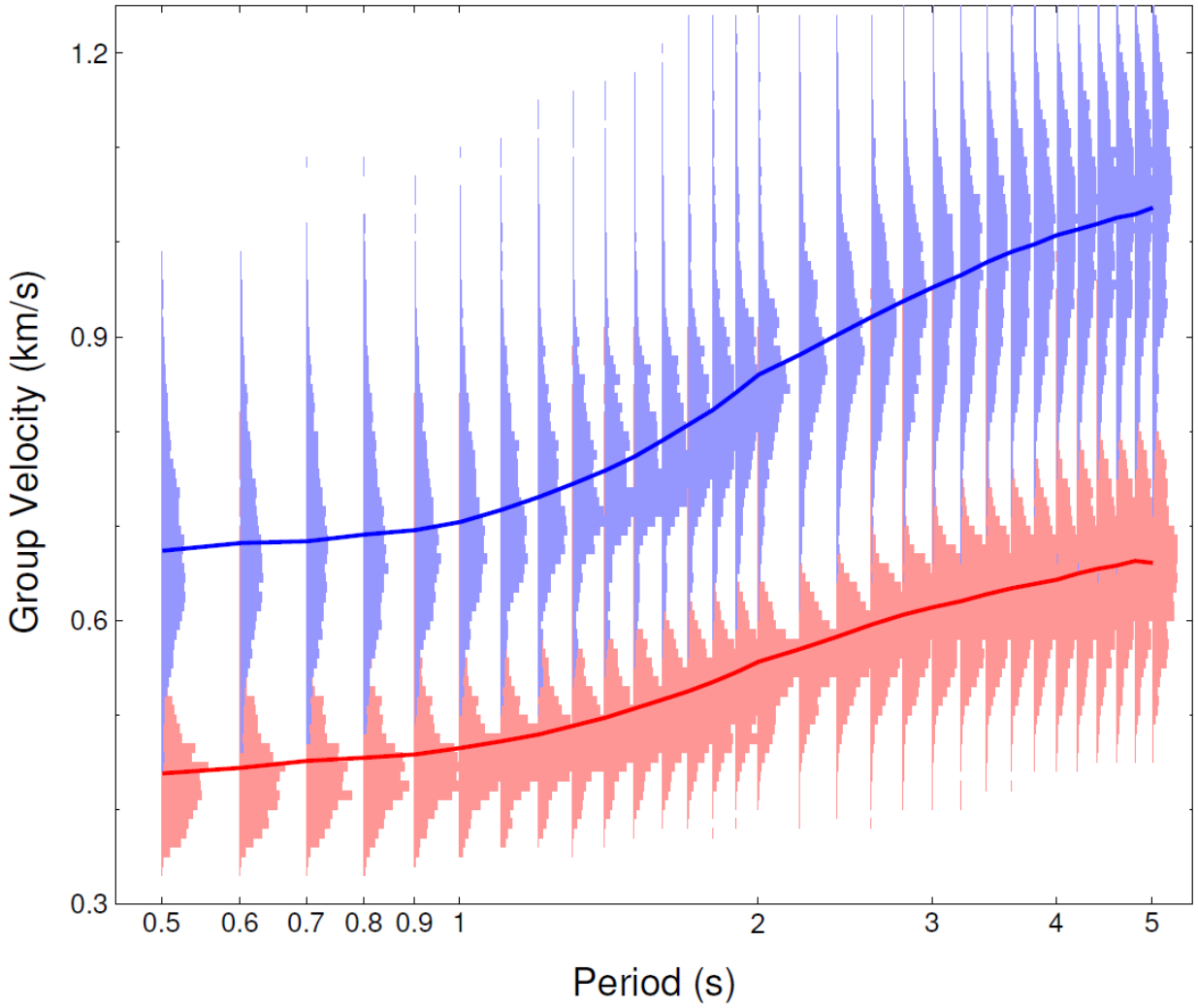


Figure 5. Compilation of all used dispersion picks. The histograms indicate distributions of the fundamental mode (in light red) and first overtone (in light blue) group velocity picks for different periods. The period axis uses a logarithmic scale. Lines in red and blue indicate the mean group velocities for fundamental mode and first overtone, respectively.

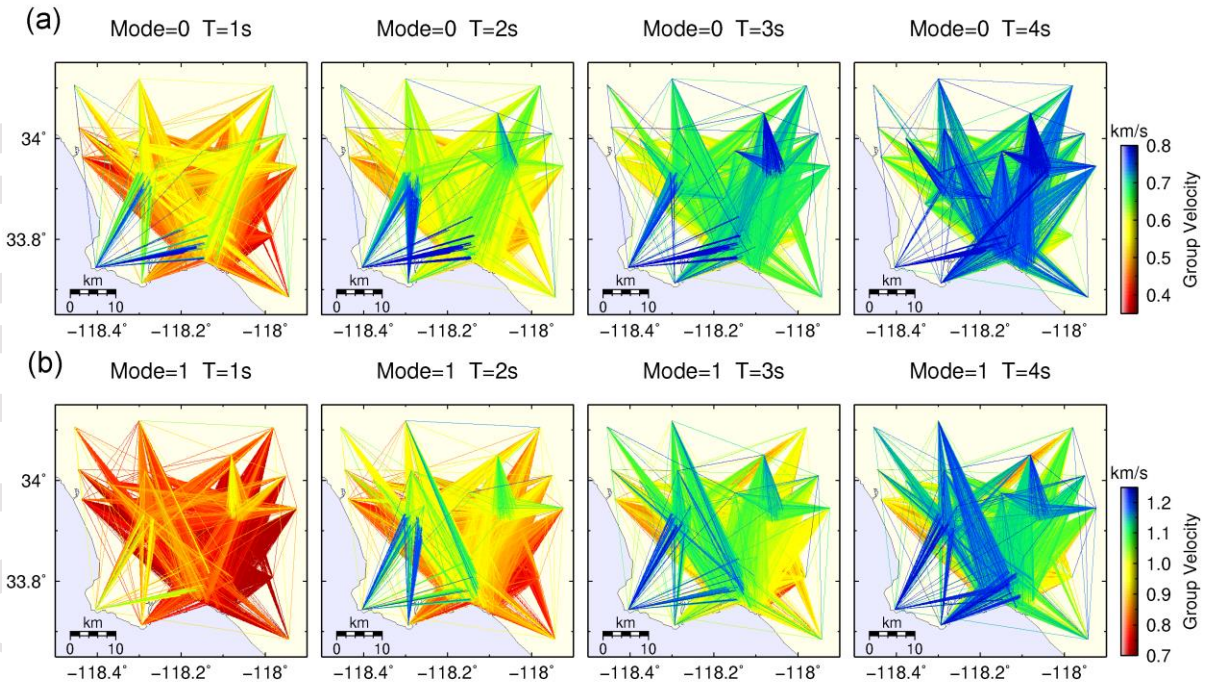


Figure 6. Ray coverage of (a) the fundamental mode and (b) the first overtone Rayleigh wave group velocity measurements for periods 1-4 s.

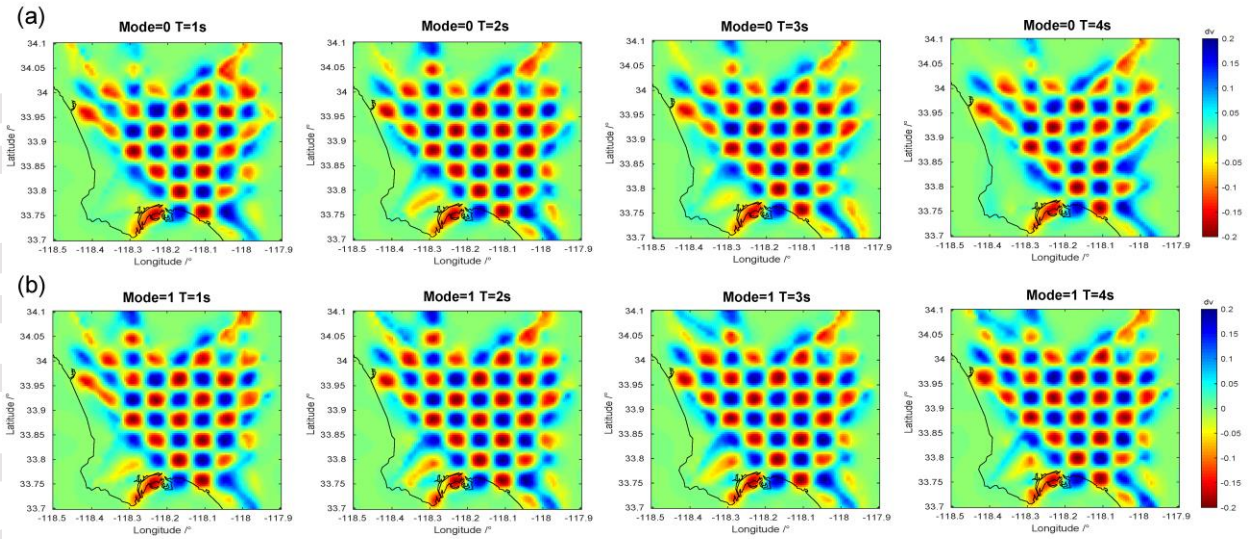


Figure 7. Checkboard resolution maps of (a) the fundamental mode and (b) the first overtone Rayleigh wave group velocity tomography for periods 1-4 s.

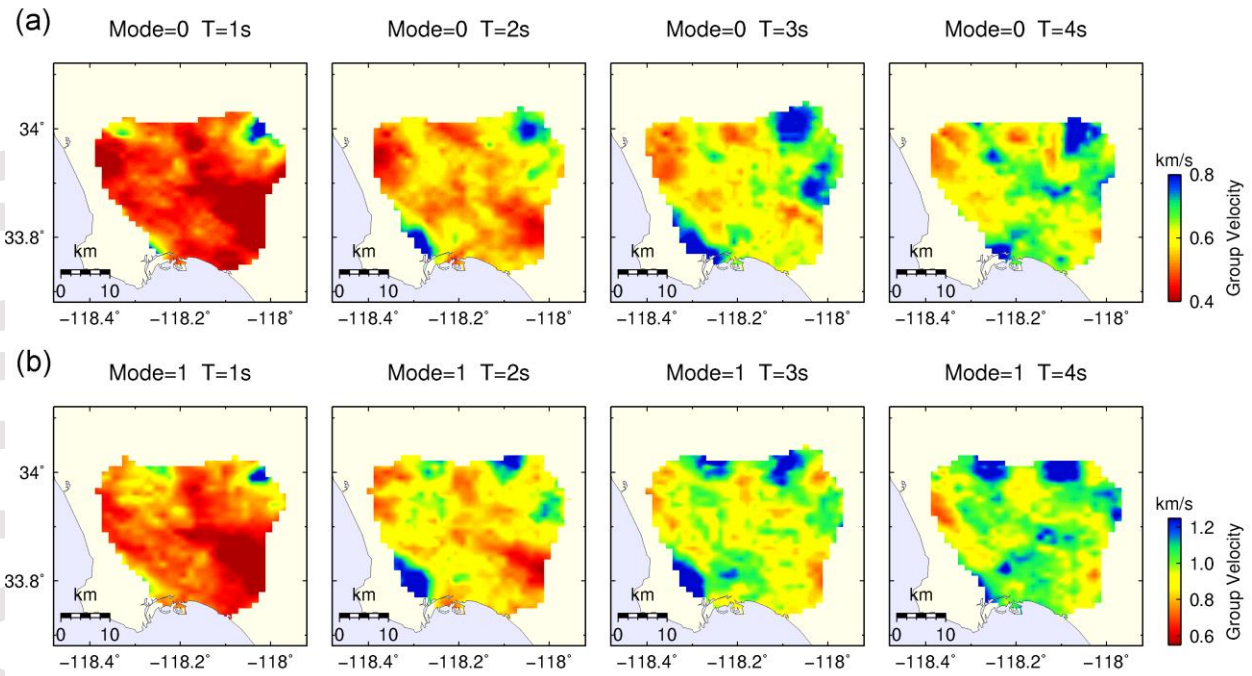


Figure 8. Group velocity maps of (a) the fundamental mode and (b) the first overtone Rayleigh wave group velocity tomography for periods 1-4 s.

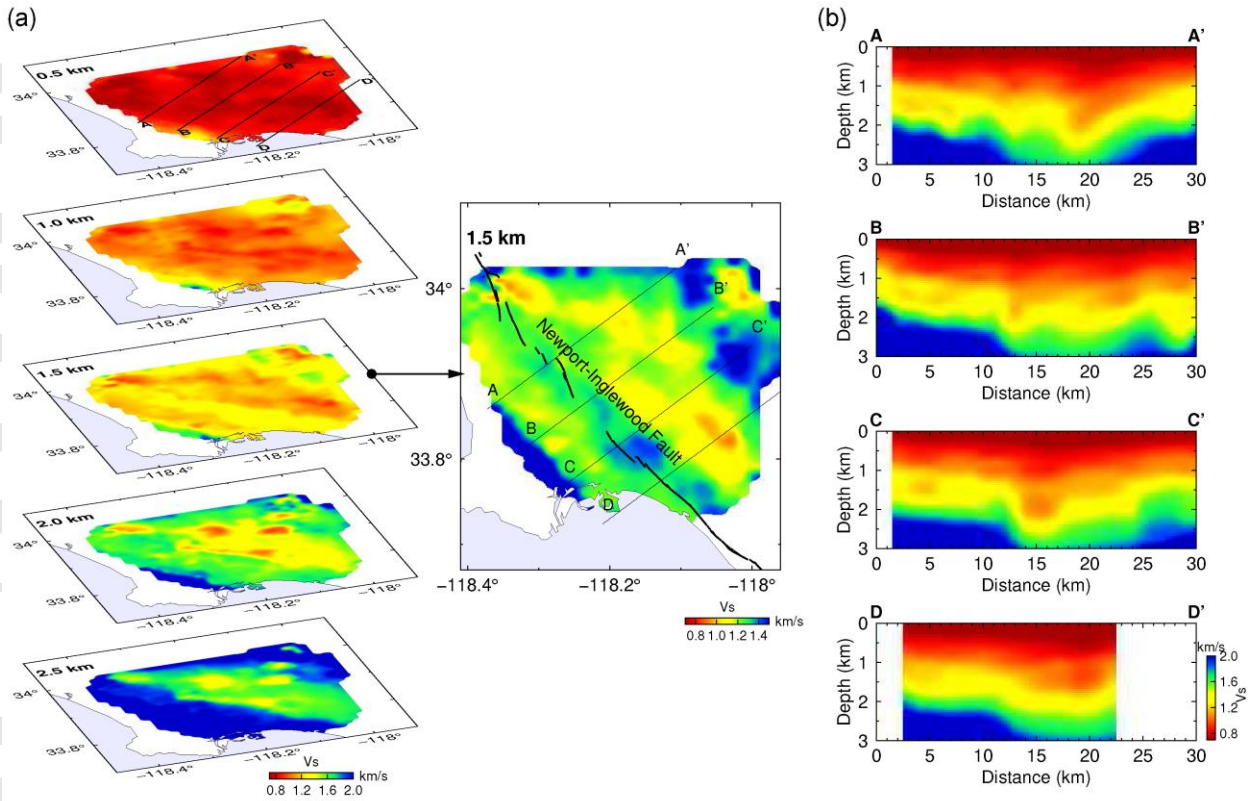


Figure 9. The inverted shear wave velocity model of the LA Basin. (a) Horizontal slides of the Vs model at depths of 0.5, 1.0, 1.5, 2.0, and 2.5 km. The center map is a rescaled version of the 1.5 km depth slice to illuminate the NW-SE trending high velocity zone. Black solid lines indicate the Newport-Inglewood faults. (b) Vertical cross sections of the Vs model along 4 profiles, AA', BB', CC', and DD' in (a).

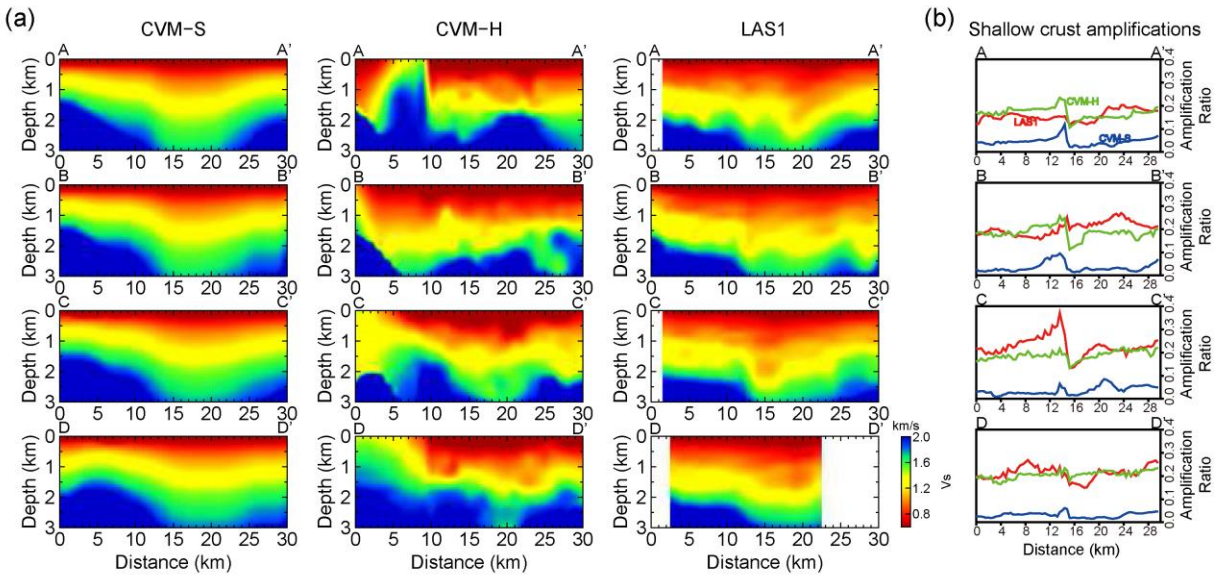


Figure 10. Comparison of our inverted LAS1 Vs model with CVM-S and CVM-H models. (a) Model comparisons for 4 cross sections in Fig. 9. (b) Prediction of strong motion amplifications along the 4 cross sections using these Vs models. The red, blue and green lines indicate synthetics from the LAS1, CVM-S and CVM-H Vs models, respectively. The gapped Vs for the LAS1 cross sections are filled by replicating existing Vs at the edges to the ends of cross sections, where the amplification fluctuations are due to the reflections and scatterings in the 2D simulations.

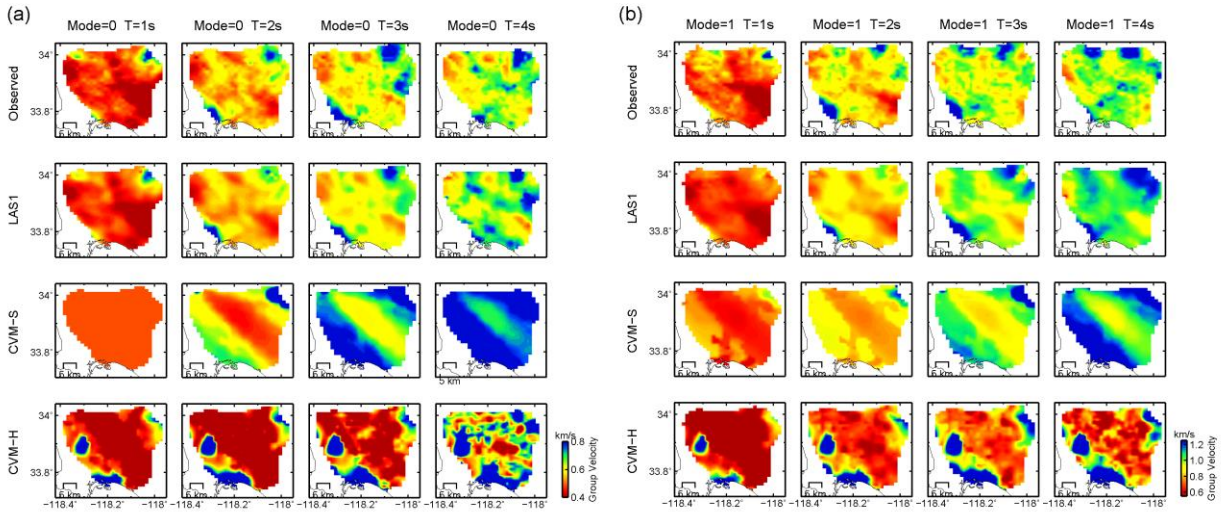


Figure 11. Comparison of (a) the fundamental mode and (b) the first overtone group velocity fittings between our Vs model and SCEC CVM Vs models for periods 1-4s. Rows from top to bottom are the observed group velocity maps from tomography, synthetic group velocity maps from the LAS1 Vs model, synthetics from the CVM-S model, and synthetics from the CVM-H model, respectively.

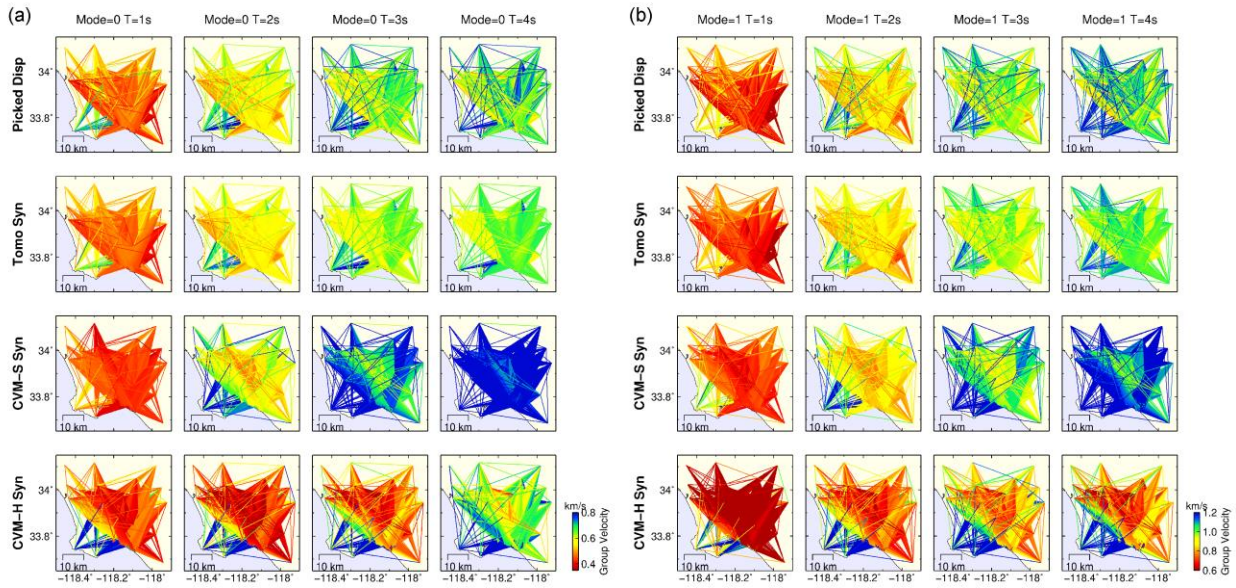
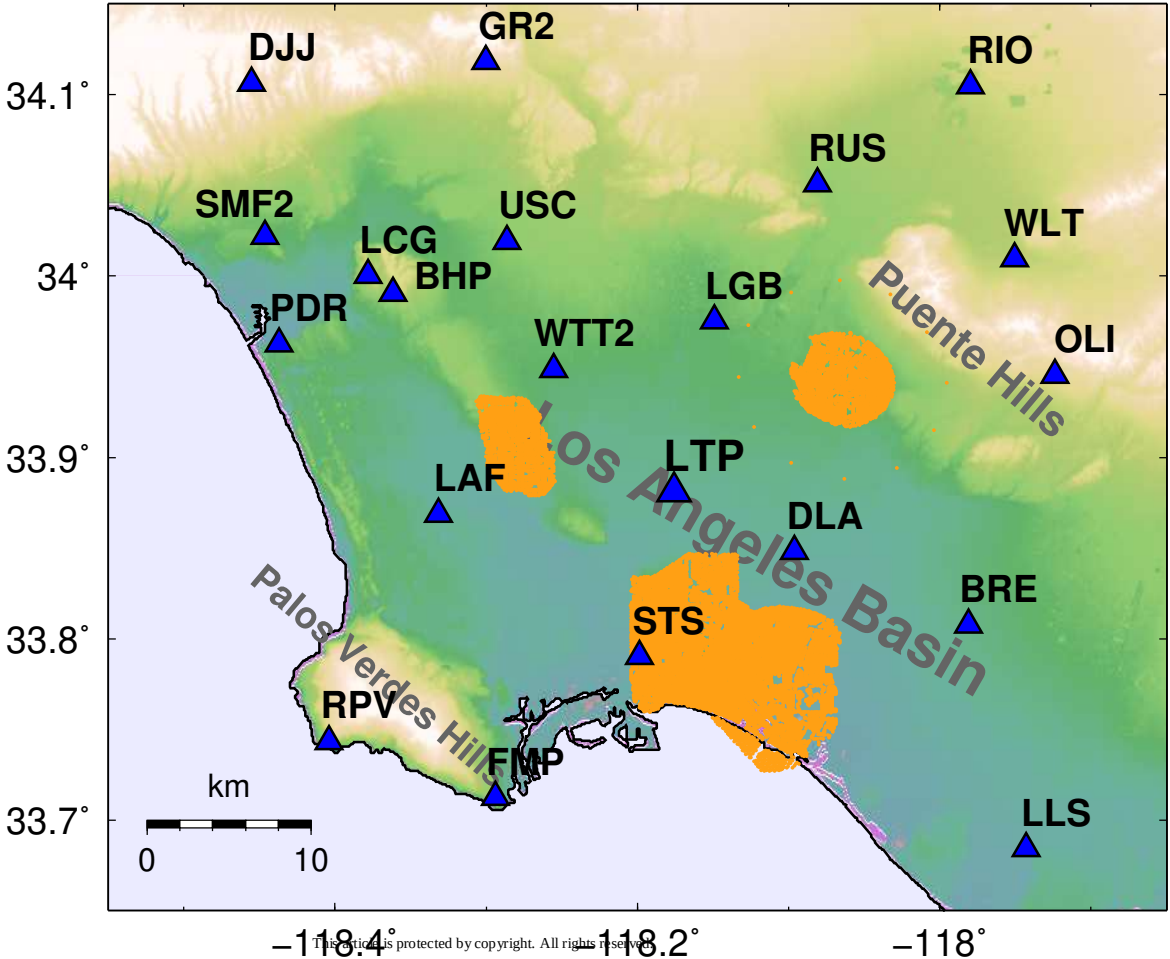
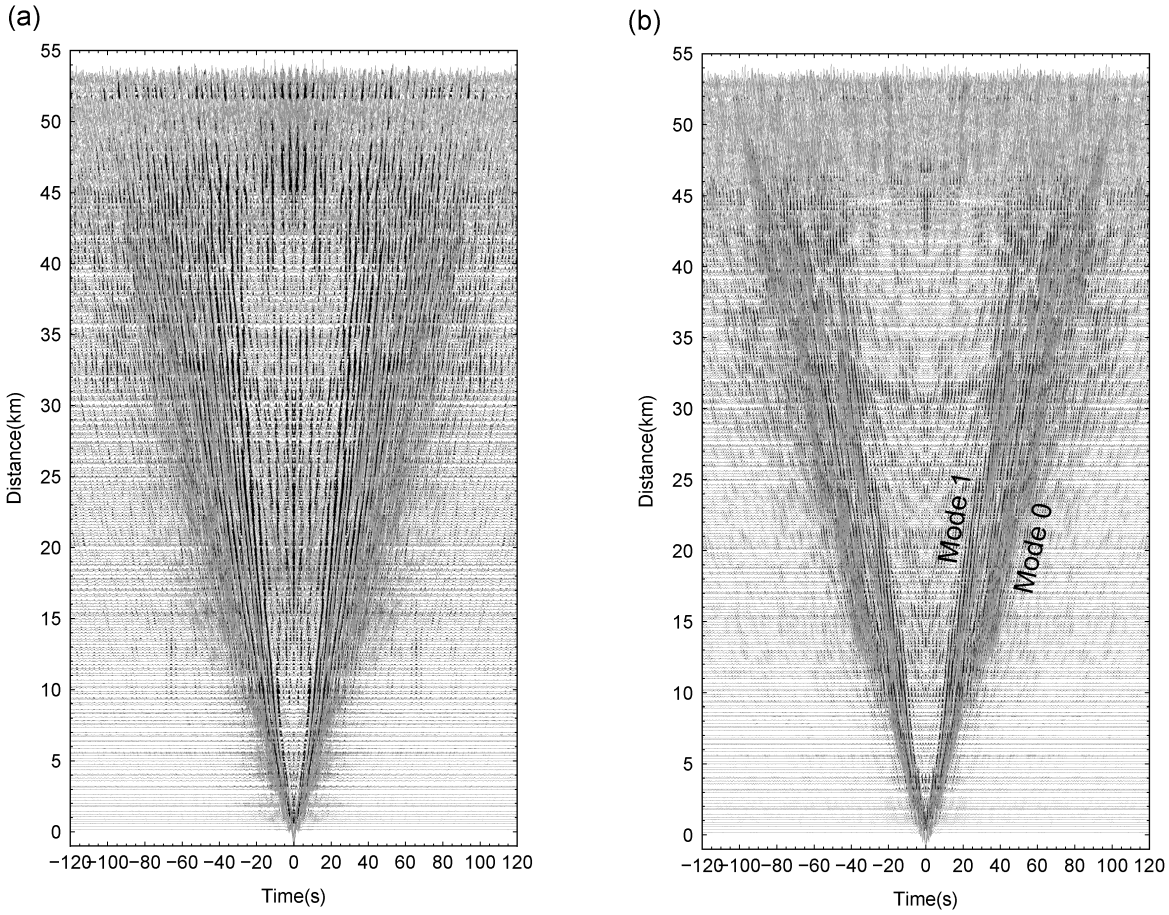
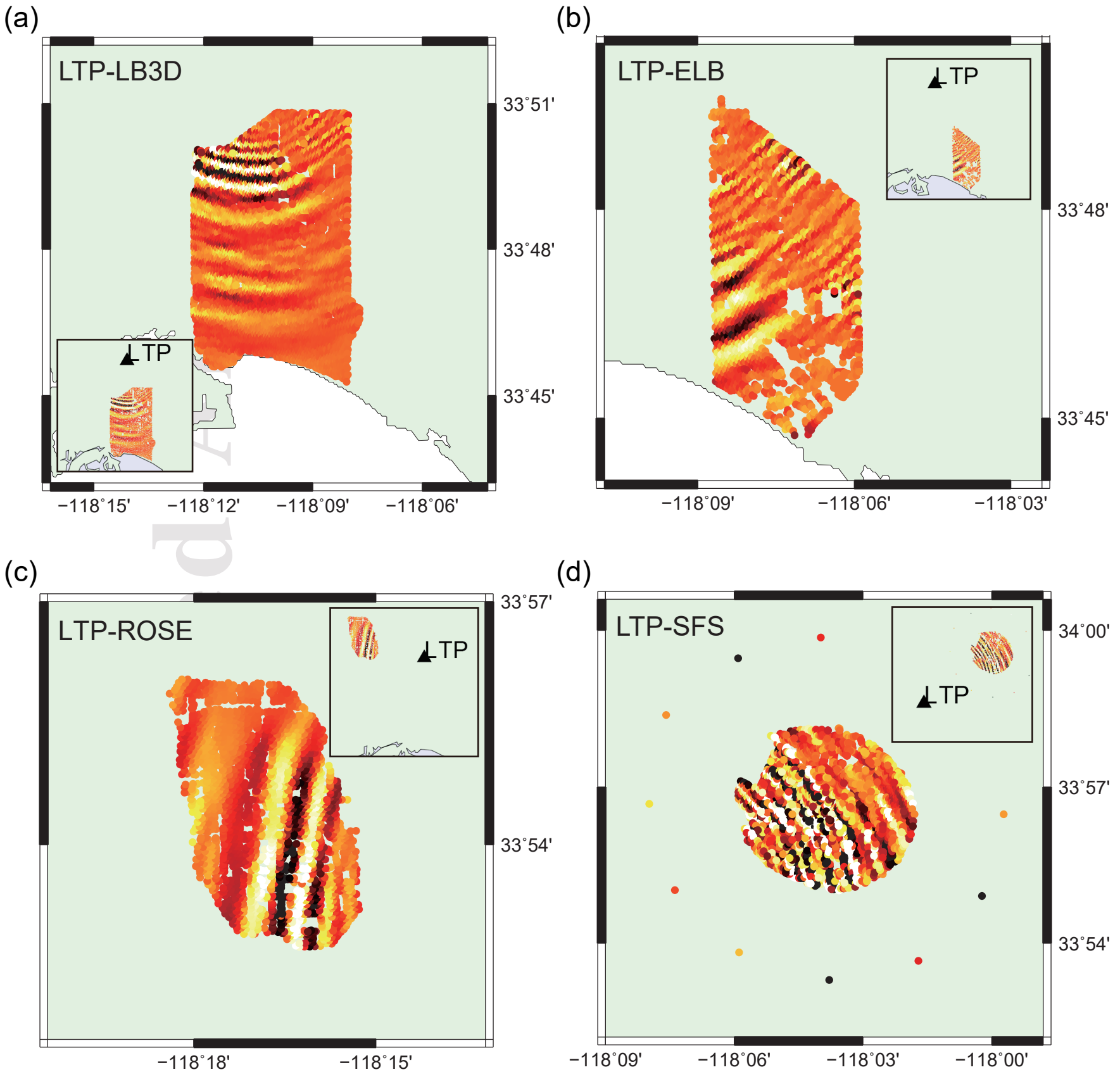
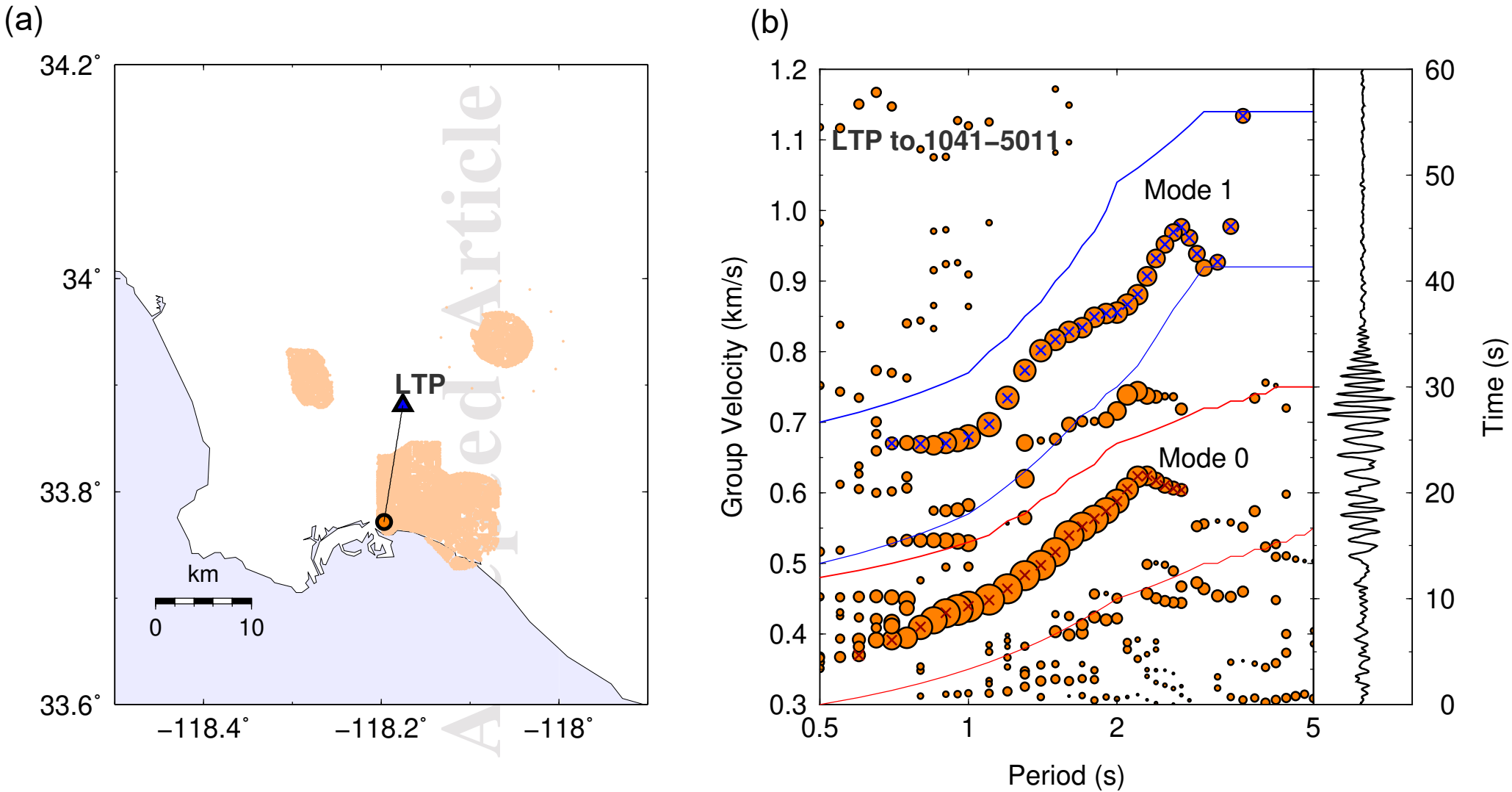


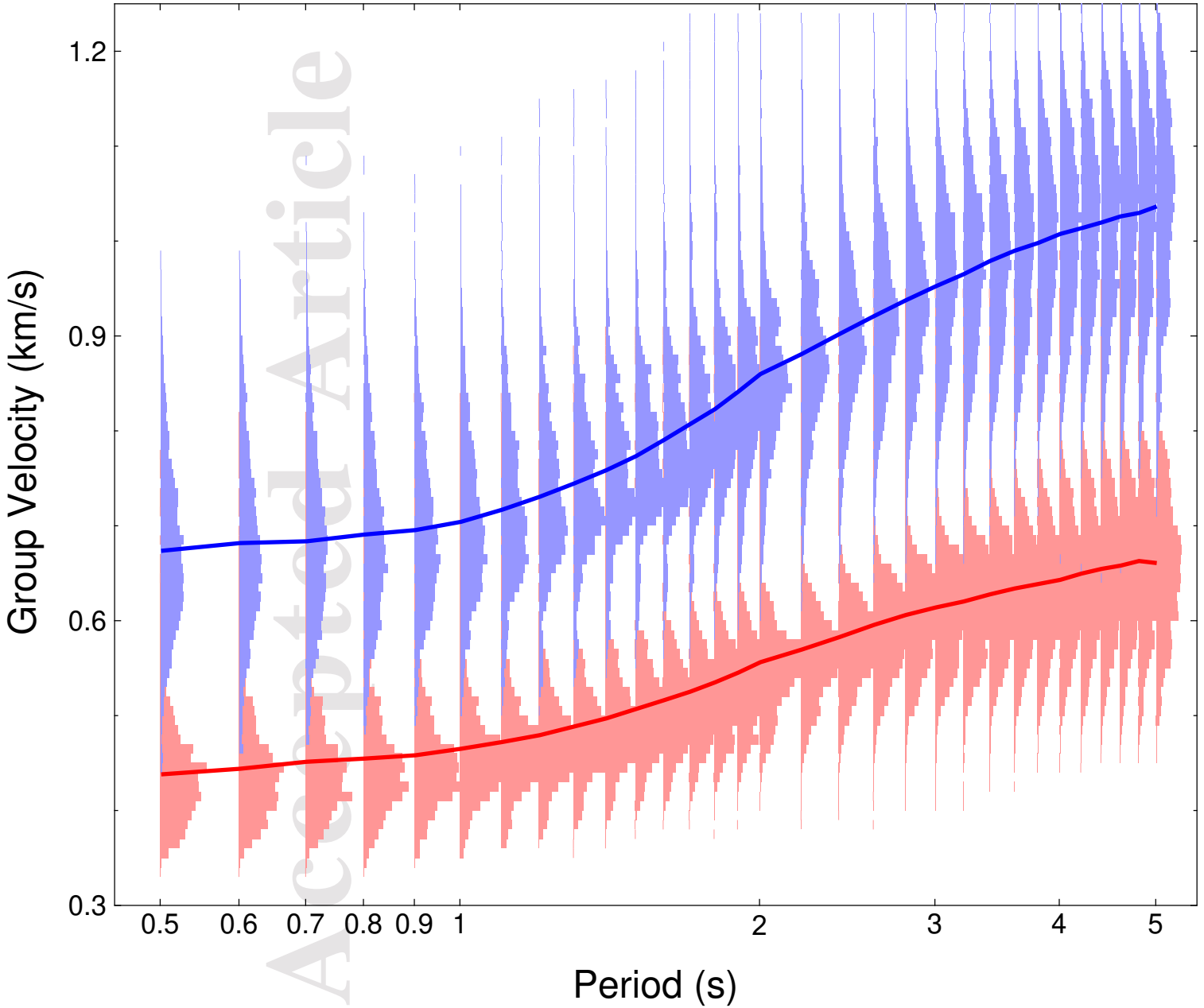
Figure 12. Comparison of (a) the fundamental mode and (b) the first overtone dispersion data fittings between our tomographic model and SCEC CVM Vs models for periods 1-4s. Rows from top to bottom are the directly picked dispersion data, synthetic dispersions of rays from our tomographic group velocity model, synthetics from the CVM-S model, and synthetics from the CVM-H model, respectively.

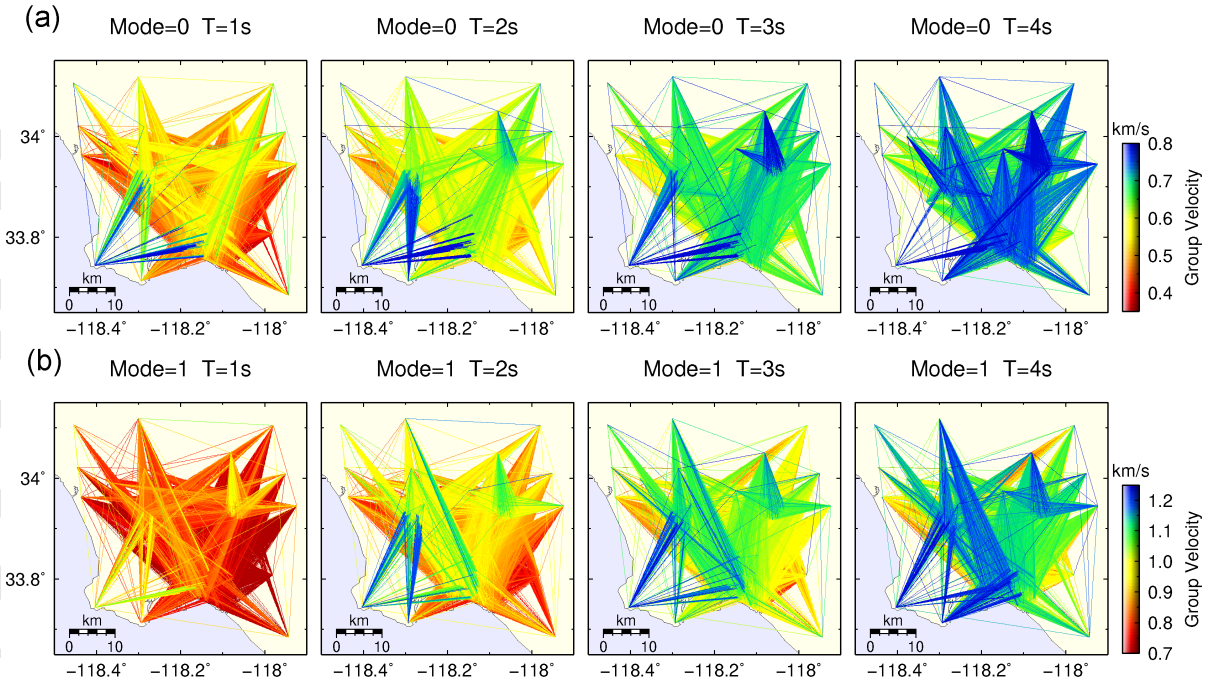


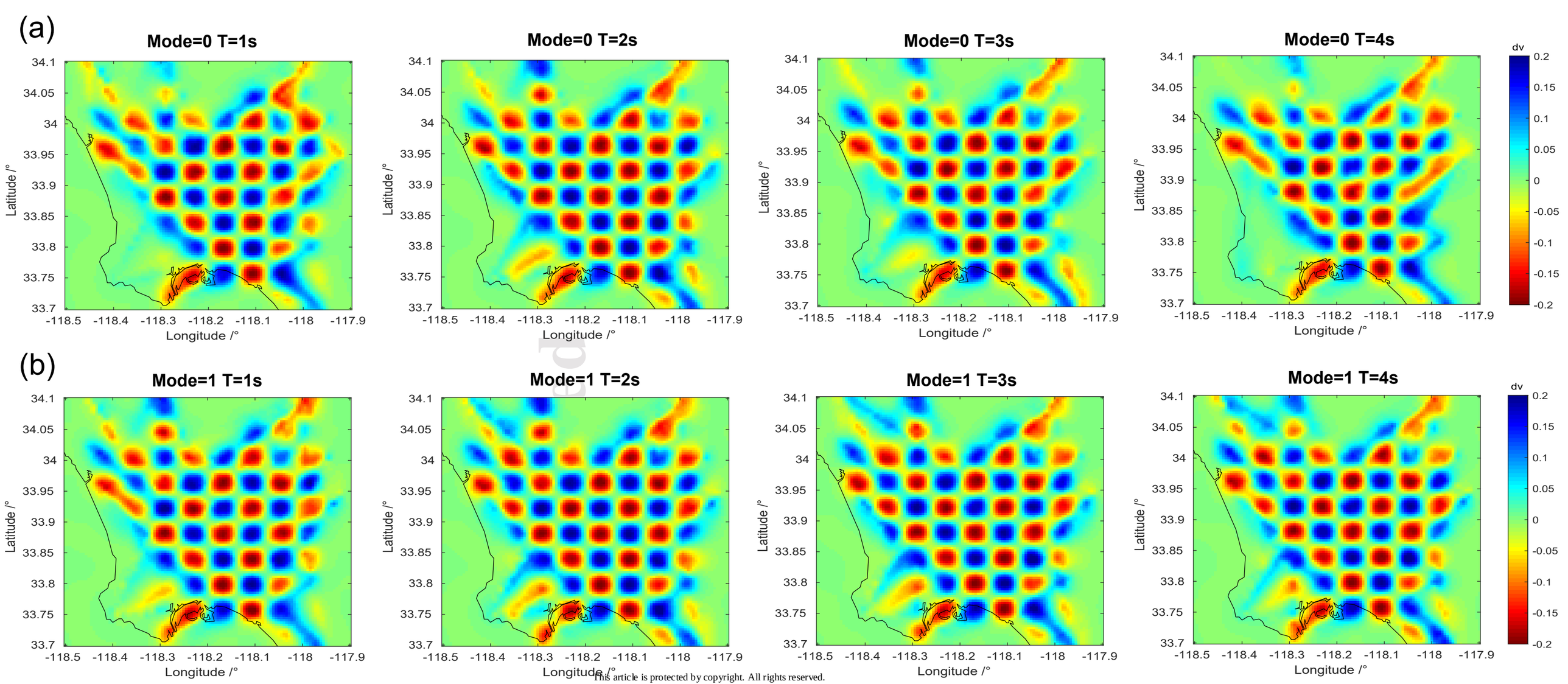


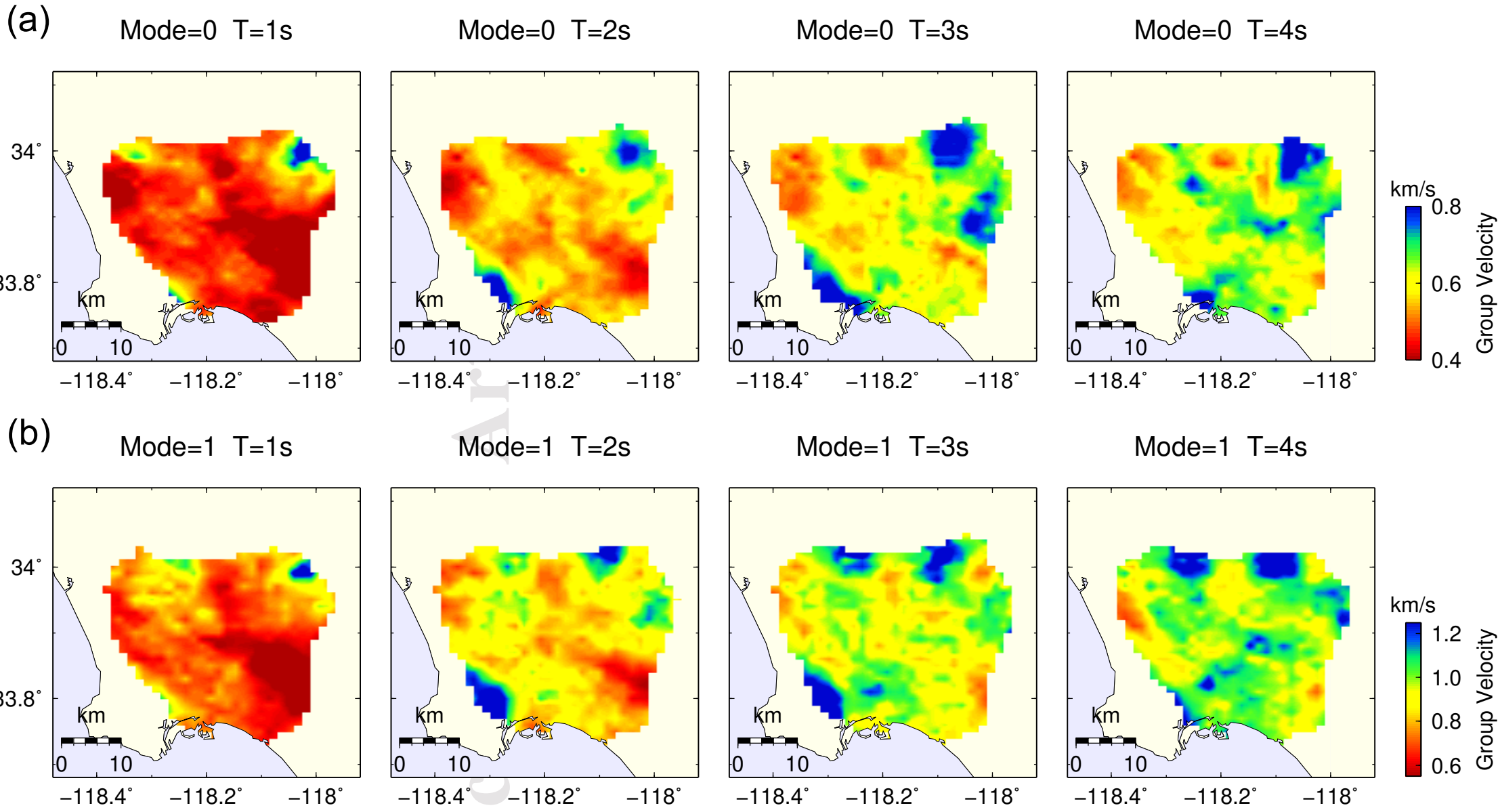











(a)
Mode=0 T=1s
Mode=0 T=2s
Mode=0 T=3s
Mode=0 T=4s
(b)
Mode=1 T=1s
Mode=1 T=2s
Mode=1 T=3s
Mode=1 T=4s
34˚
33.8˚
km
0 10
−118.4˚ −118.2˚ −118˚
km/s
0.8
0.6
0.4
Group Velocity
1.2
1.0
0.8
0.6